TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Statistics for Knowledge Graph Modelling

Marcin Pietrasik
University of Alberta

June 24, 2022 // Dresden, Germany

Motivation

- A couple has two children. At least one of the children is a boy. What is the probability of the other child also being a boy?

Motivation

- A couple has two children. At least one of the children is a boy. What is the probability of the other child also being a boy?
  - $^{1}/_{3}$

Motivation

- A couple has two children. At least one of the children is a boy. What is the probability of the other child also being a boy?
  - $^1/_3$

First Child	Second Child
Boy	Boy
Boy	Girl
Girl	Boy
Girl	Girl

Motivation

- A couple has two children. At least one of the children is a boy. What is the probability of the other child also being a boy?
  - ${}^{1}/_{3}$

First Child	Second Child
Boy	Boy
Boy	Girl
Girl	Boy
Girl	Girl

Motivation

- A couple has two children. At least one of the children is a boy. What is the probability of the other child also being a boy?
  - $^{1}/_{3}$

First Child	Second Child
Boy	Boy
Boy	Girl
Girl	Boy

Motivation

- A couple has two children. At least one of the children is a boy. What is the probability of the other child also being a boy?
  - $^1/_3$

First Child	Second Child
Boy	Boy
Boy	Girl
Girl	Boy

- Boy or Girl Paradox, Mr. Smith's children problem, etc.

Outline

- Probability vocabulary
- Chain and Bayes’ rules
- Gibbs sampling
- My work

Running example

		Raining	
		1	0
Umbrella	1		
	0		

- Model the relationship between it raining outside (R) and taking an umbrella to work (U).

Running example

		Raining	
		1	0
Umbrella	1		
	0		

- Model the relationship between it raining outside (R) and taking an umbrella to work (U).
- R and U are random variables and can take on two values:
  - $R = 1$ raining $\qquad$ $R = 0$ not raining
  - $U = 1$ umbrella $\qquad$ $U = 0$ no umbrella

Joint distribution

		Raining	
		1	0
Umbrella	1		
	0		

- The joint distribution $P(R, U)$ is the probability of two or more random variables occurring together.

Joint distribution

		Raining	
		1	0
Umbrella	1	0.15	0.1
	0	0.05	0.7

- The joint distribution $P(R, U)$ is the probability of two or more random variables observed together.
- The probability of it raining and taking an umbrella to work is 0.15 or 15%.
  - $P(R = 1, U = 1) = 0.15$

Marginal distribution

		Raining: 1	Raining: 0
Umbrella	1	0.15	0.1
	0	0.05	0.7

- The marginal distribution $P(R)$ is the probability of observing a random variable irrespective of all other variables.
  - Marginal because they are calculated by summing over the rows and columns of the table.

Marginal distribution

		Raining		
		1	0	
Umbrella	1	0.15	0.1	0.25
	0	0.05	0.7	0.75
		0.2	0.8	

- The marginal distribution $P(R)$ is the probability of observing a random variable irrespective of all other variables.
  - Marginal because they are calculated by summing over the rows and columns of the table.
- The probability of taking an umbrella to work, regardless of whether it's raining, is 0.25 or 25%.
  - $P(U = 1) = 0.25$

Conditional distribution

		Raining		
		1	0	
Umbrella	1	0.15	0.1	0.25
	0	0.05	0.7	0.75
		0.2	0.8	

- The conditional distribution $P(U \mid R)$ is the probability of observing a random variable given the occurrence of another variable.
  - Calculated as the joint distribution over the marginal.

Conditional distribution

		Raining	
		1	0
Umbrella	1	0.15	
	0	0.05	
		0.2	

- The conditional distribution $P(U \mid R)$ is the probability of observing a random variable given the occurrence of another variable.
  - Calculated as the joint distribution over the marginal.
- The probability of taking an umbrella to work given that it is raining is 0.75 or 75%.
  - $P(U = 1 \mid R = 1) = \frac{P(U=1,R=1)}{P(R=1)} = \frac{0.15}{0.2} = 0.75$

Chain and Bayes' rules

- **Chain rule:** $P(A, B) = P(A \mid B)P(B)$
  - If we want to find the probability that it is raining and an umbrella is taken to work, first find the probability it is raining and multiply that by the probability of taking an umbrella given that it's raining.

Chain and Bayes' rules

- **Chain rule:** $P(A, B) = P(A \mid B)P(B)$
  - If we want to find the probability that it is raining and an umbrella is taken to work, first find the probability it is raining and multiply that by the probability of taking an umbrella given that it's raining.
- **Bayes' rule**: $P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$
  - Posterior $\boldsymbol{P(A \mid B)}$
  - Likelihood $\boldsymbol{P(B \mid A)}$
  - Prior $\boldsymbol{P(A)}$
  - Evidence $\boldsymbol{P(B)}$

Chain and Bayes' rules

- **Chain rule:** $P(A, B) = P(A \mid B)P(B)$
  - If we want to find the probability that it is raining and an umbrella is taken to work, first find the probability it is raining and multiply that by the probability of taking an umbrella given that it's raining.
- **Bayes' rule**: $P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$
  - Posterior $\boldsymbol{P(A \mid B)}$
  - Likelihood $\boldsymbol{P(B \mid A)}$
  - Prior $\boldsymbol{P(A)}$
  - Evidence $\boldsymbol{P(B)}$
- $P(parameters \mid data) = \frac{P(data \mid parameters)P(parameters)}{P(data)}$

Chain and Bayes' rules

- **Chain rule:** $P(A, B) = P(A \mid B)P(B)$
  - If we want to find the probability that it will rain and an umbrella is taken to work, first find the probability it will rain and multiply that by the probability of taking an umbrella given that it's raining.
- **Bayes' rule**: $P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$
  - Posterior $\boldsymbol{P(A \mid B)}$
  - Likelihood $\boldsymbol{P(B \mid A)}$
  - Prior $\boldsymbol{P(A)}$
  - Evidence $\boldsymbol{P(B)}$
- $P(parameters \mid data) = \frac{P(data \mid parameters)P(parameters)}{P(data)} \propto P(data \mid parameters)P(parameters)$

Bayes’ rule - example

- A couple has two children. At least one of the children is a boy. What is the probability of the other child also being a boy?

Bayes' rule - example

- A couple has two children. At least one of the children is a boy. What is the probability of the other child also being a boy?
  - We want to find the conditional $P(\{boy, boy\} \mid \{boy\})$

Bayes' rule - example

- A couple has two children. At least one of the children is a boy. What is the probability of the other child also being a boy?
  - We want to find the conditional $P(\{boy, boy\} \mid \{boy\})$
- We can apply Bayes' rule!
  - Likelihood $P(\{boy\} \mid \{boy, boy\}) = 1$
  - Prior $P(\{boy, boy\}) = \frac{1}{4}$
  - Evidence $P(\{boy\}) = \frac{3}{4}$

First Child	Second Child
Boy	Boy
Boy	Girl
Girl	Boy
Girl	Girl

Bayes' rule - example

- A couple has two children. At least one of the children is a boy. What is the probability of the other child also being a boy?
  - We want to find the conditional $P(\{boy, boy\} \mid \{boy\})$
- We can apply Bayes' rule!
  - Likelihood $P(\{boy\} \mid \{boy, boy\}) = 1$
  - Prior $P(\{boy, boy\}) = \frac{1}{4}$
  - Evidence $P(\{boy\}) = \frac{3}{4}$
- Posterior calculation:
  - $P(\{boy, boy\} \mid \{boy\}) = \frac{P(\{boy\} \mid \{boy,boy\})P(\{boy,boy\})}{P(\{boy\})} = \frac{1 * {}^1/_4}{{}^3/_4} = \frac{1}{3}$

First Child	Second Child
Boy	Boy
Boy	Girl
Girl	Boy
Girl	Girl

Gibbs sampling

- **Markov chain Monte Carlo** technique for approximating a joint distribution when it's difficult to sample directly.
    - Only possible when conditional distributions are easy to sample from.

Gibbs sampling

- **Markov chain Monte Carlo** technique for approximating a joint distribution when it's difficult to sample directly.
  - Only possible when conditional distributions are easy to sample from.
- General idea: iteratively sample from conditionals to draw samples from the joint distribution.

Gibbs sampling

- **Markov chain Monte Carlo** technique for approximating a joint distribution when it's difficult to sample directly.
  - Only possible when conditional distributions are easy to sample from.
- General idea: iteratively sample from conditionals to draw samples from the joint distribution.
- Algorithm for sampling $P(R, U)$:

1. Initialize R and U with some values in the support
2. For i iterations
   1. Obtain new R from $P(R \mid U)$
   2. Obtain new U from $P(U \mid R)$

Gibbs sampling - example

$P(R, U)$

		Raining 1	Raining 0
Umbrella	1	0.15	0.1
Umbrella	0	0.05	0.7

- Going back to the rain and umbrella example, say we want to sample from $P(R, U)$ but doing so directly is hard.

Gibbs sampling - example

$P(R, U)$

Umbrella \ Raining	1	0
1	0.15	0.1
0	0.05	0.7

$P(R \mid U)$

Umbrella \ Raining	1	0
1		
0		

$P(U \mid R)$

Umbrella \ Raining	1	0
1		
0		

- Going back to the rain and umbrella example, say we want to sample from $P(R, U)$ but doing so directly is hard.
- However, sampling from the conditionals is easy (easier).

Gibbs sampling - example

$P(R, U)$

Umbrella \ Raining	1	0	
1	0.15	0.1	0.25
0			

$P(R \mid U)$

Umbrella \ Raining	1	0
1	0.6	0.4
0		

$P(U \mid R)$

Umbrella \ Raining	1	0
1		
0		

- Going back to the rain and umbrella example, say we want to sample from $P(R, U)$ but doing so directly is hard.
- However, sampling from the conditionals is easy (easier).

Gibbs sampling - example

$P(R, U)$

Umbrella \ Raining	1	0	
1			
0	0.05	0.7	0.75

$P(R \mid U)$

Umbrella \ Raining	1	0
1	0.6	0.4
0	0.07	0.93

$P(U \mid R)$

Umbrella \ Raining	1	0
1		
0		

- Going back to the rain and umbrella example, say we want to sample from $P(R, U)$ but doing so directly is hard.
- However, sampling from the conditionals is easy (easier).

Gibbs sampling - example

$P(R, U)$

		Raining 1	Raining 0
Umbrella	1	0.15	0.1
Umbrella	0	0.05	0.7

$P(R \mid U)$

		Raining 1	Raining 0
Umbrella	1	0.6	0.4
Umbrella	0	0.07	0.93

$P(U \mid R)$

		Raining 1	Raining 0
Umbrella	1	0.75	0.12
Umbrella	0	0.25	0.88

- Going back to the rain and umbrella example, say we want to sample from $P(R, U)$ but doing so directly is hard.
- However, sampling from the conditionals is easy (easier).

Gibbs sampling - example

$P(R, U)$

Umbrella \ Raining	1	0
1	0.15	0.1
0	0.05	0.7

$P(R \mid U)$

Umbrella \ Raining	1	0
1	0.6	0.4
0	0.07	0.93

$$P(R = 1 \mid U = 1) = 0.6$$
$$P(R = 0 \mid U = 1) = 0.4$$
$$P(R = 0 \mid U = 0) = 0.93$$
$$P(R = 1 \mid U = 0) = 0.07$$

$P(U \mid R)$

Umbrella \ Raining	1	0
1	0.75	0.12
0	0.25	0.88

$$P(U = 1 \mid R = 1) = 0.75$$
$$P(U = 0 \mid R = 1) = 0.25$$
$$P(U = 0 \mid R = 0) = 0.88$$
$$P(U = 1 \mid R = 0) = 0.12$$

- Going back to the rain and umbrella example, say we want to sample from $P(R, U)$ but doing so directly is hard.
- However, sampling from the conditionals is easy (easier).

Gibbs sampling - example

$P(R, U)$		Raining 1	Raining 0
Umbrella	1	0	0
	0	0	0

$$P(R = 1 \mid U = 1) = 0.6$$
$$P(R = 0 \mid U = 1) = 0.4$$
$$P(R = 0 \mid U = 0) = 0.93$$
$$P(R = 1 \mid U = 0) = 0.07$$

$$P(U = 1 \mid R = 1) = 0.75$$
$$P(U = 0 \mid R = 1) = 0.25$$
$$P(U = 0 \mid R = 0) = 0.88$$
$$P(U = 1 \mid R = 0) = 0.12$$

1. Initialize R and U with some values in the support
2. For *i* iterations
    1. Obtain new R from $P(R \mid U)$
    2. Obtain new U from $P(U \mid R)$

Gibbs sampling - example

		Raining	
$P(R,U)$		1	0
Umbrella	1	1	0
	0	0	0

$$P(R = 1 \mid U = 1) = 0.6$$
$$P(R = 0 \mid U = 1) = 0.4$$
$$P(R = 0 \mid U = 0) = 0.93$$
$$P(R = 1 \mid U = 0) = 0.07$$

$$P(U = 1 \mid R = 1) = 0.75$$
$$P(U = 0 \mid R = 1) = 0.25$$
$$P(U = 0 \mid R = 0) = 0.88$$
$$P(U = 1 \mid R = 0) = 0.12$$

1. **Initialize *R* and *U* with some values in the support**
2. For *i* iterations
   1. Obtain new R from $P(R \mid U)$
   2. Obtain new U from $P(U \mid R)$

Gibbs sampling - example

Raining

$P(R, U)$	1	0
Umbrella 1	1	0
Umbrella 0	0	0

$\boldsymbol{P(R = 1 \mid U = 1) = 0.6}$
$\boldsymbol{P(R = 0 \mid U = 1) = 0.4}$
$P(R = 0 \mid U = 0) = 0.93$
$P(R = 1 \mid U = 0) = 0.07$

$P(U = 1 \mid R = 1) = 0.75$
$P(U = 0 \mid R = 1) = 0.25$
$P(U = 0 \mid R = 0) = 0.88$
$P(U = 1 \mid R = 0) = 0.12$

1. Initialize R and U with some values in the support
2. For *i* iterations
    1. **Obtain new $\boldsymbol{R}$ from $\boldsymbol{P(R \mid U)}$**
    2. Obtain new U from $P(U \mid R)$

Gibbs sampling - example

		Raining	
	$P(R, U)$	1	0
Umbrella	1	1	1
	0	0	0

$\boldsymbol{P(R = 1 \mid U = 1) = 0.6}$
$\boldsymbol{P(R = 0 \mid U = 1) = 0.4}$
$P(R = 0 \mid U = 0) = 0.93$
$P(R = 1 \mid U = 0) = 0.07$

$P(U = 1 \mid R = 1) = 0.75$
$P(U = 0 \mid R = 1) = 0.25$
$P(U = 0 \mid R = 0) = 0.88$
$P(U = 1 \mid R = 0) = 0.12$

1. Initialize R and U with some values in the support
2. For *i* iterations
   1. **Obtain new $\boldsymbol{R}$ from $\boldsymbol{P(R \mid U)}$**
   2. Obtain new U from $P(U \mid R)$

Gibbs sampling - example

$P(R, U)$	Raining: 1	Raining: 0
Umbrella: 1	1	1
Umbrella: 0	0	0

$$P(R = 1 \mid U = 1) = 0.6$$
$$P(R = 0 \mid U = 1) = 0.4$$
$$P(R = 0 \mid U = 0) = 0.93$$
$$P(R = 1 \mid U = 0) = 0.07$$

$$P(U = 1 \mid R = 1) = 0.75$$
$$P(U = 0 \mid R = 1) = 0.25$$
$$\boldsymbol{P(U = 0 \mid R = 0) = 0.88}$$
$$\boldsymbol{P(U = 1 \mid R = 0) = 0.12}$$

1. Initialize R and U with some values in the support
2. For *i* iterations
   1. Obtain new R from $P(R \mid U)$
   2. **Obtain new U from $P(U \mid R)$**

Gibbs sampling - example

Raining

$P(R, U)$		1	0
Umbrella	1	1	1
	0	0	1

$$P(R = 1 \mid U = 1) = 0.6$$
$$P(R = 0 \mid U = 1) = 0.4$$
$$P(R = 0 \mid U = 0) = 0.93$$
$$P(R = 1 \mid U = 0) = 0.07$$

$$P(U = 1 \mid R = 1) = 0.75$$
$$P(U = 0 \mid R = 1) = 0.25$$
$$\boldsymbol{P(U = 0 \mid R = 0) = 0.88}$$
$$\boldsymbol{P(U = 1 \mid R = 0) = 0.12}$$

1. Initialize R and U with some values in the support
2. For *i* iterations
    1. Obtain new R from $P(R \mid U)$
    2. **Obtain new U from $P(U \mid R)$**

Gibbs sampling - example

		Raining	
$P(R, U)$		1	0
Umbrella	1	1	1
	0	0	1

$P(R = 1 \mid U = 1) = 0.6$
$P(R = 0 \mid U = 1) = 0.4$
$\boldsymbol{P(R = 0 \mid U = 0) = 0.93}$
$\boldsymbol{P(R = 1 \mid U = 0) = 0.07}$

$P(U = 1 \mid R = 1) = 0.75$
$P(U = 0 \mid R = 1) = 0.25$
$P(U = 0 \mid R = 0) = 0.88$
$P(U = 1 \mid R = 0) = 0.12$

1. Initialize R and U with some values in the support
2. For *i* iterations
   1. **Obtain new $\boldsymbol{R}$ from $\boldsymbol{P(R \mid U)}$**
   2. Obtain new U from $P(U \mid R)$

Gibbs sampling - example

$P(R,U)$	Raining: 1	Raining: 0
Umbrella: 1	1	1
Umbrella: 0	0	2

$P(R = 1 \mid U = 1) = 0.6$
$P(R = 0 \mid U = 1) = 0.4$
$\boldsymbol{P(R = 0 \mid U = 0) = 0.93}$
$\boldsymbol{P(R = 1 \mid U = 0) = 0.07}$

$P(U = 1 \mid R = 1) = 0.75$
$P(U = 0 \mid R = 1) = 0.25$
$P(U = 0 \mid R = 0) = 0.88$
$P(U = 1 \mid R = 0) = 0.12$

1. Initialize R and U with some values in the support
2. For *i* iterations
   1. **Obtain new $\boldsymbol{R}$ from $\boldsymbol{P(R \mid U)}$**
   2. Obtain new U from $P(U \mid R)$

Gibbs sampling - example

$P(R,U)$		Raining 1	Raining 0
Umbrella	1	1	1
	0	0	2

$$P(R = 1 \mid U = 1) = 0.6$$
$$P(R = 0 \mid U = 1) = 0.4$$
$$P(R = 0 \mid U = 0) = 0.93$$
$$P(R = 1 \mid U = 0) = 0.07$$

$$P(U = 1 \mid R = 1) = 0.75$$
$$P(U = 0 \mid R = 1) = 0.25$$
$$\boldsymbol{P(U = 0 \mid R = 0) = 0.88}$$
$$\boldsymbol{P(U = 1 \mid R = 0) = 0.12}$$

1. Initialize R and U with some values in the support
2. For *i* iterations
   1. Obtain new R from $P(R \mid U)$
   2. **Obtain new U from $P(U \mid R)$**

Gibbs sampling - example

$P(R,U)$		Raining 1	Raining 0
Umbrella	1	1	2
Umbrella	0	0	2

$P(R = 1 \mid U = 1) = 0.6$
$P(R = 0 \mid U = 1) = 0.4$
$P(R = 0 \mid U = 0) = 0.93$
$P(R = 1 \mid U = 0) = 0.07$

$P(U = 1 \mid R = 1) = 0.75$
$P(U = 0 \mid R = 1) = 0.25$
$\boldsymbol{P(U = 0 \mid R = 0) = 0.88}$
$\boldsymbol{P(U = 1 \mid R = 0) = 0.12}$

1. Initialize R and U with some values in the support
2. For *i* iterations
   1. Obtain new R from $P(R \mid U)$
   2. **Obtain new U from $P(U \mid R)$**

Gibbs sampling - example

$P(R, U)$	Raining: 1	Raining: 0
Umbrella: 1	1	2
Umbrella: 0	0	2

$\boldsymbol{P(R = 1 \mid U = 1) = 0.6}$
$\boldsymbol{P(R = 0 \mid U = 1) = 0.4}$
$P(R = 0 \mid U = 0) = 0.93$
$P(R = 1 \mid U = 0) = 0.07$

$P(U = 1 \mid R = 1) = 0.75$
$P(U = 0 \mid R = 1) = 0.25$
$P(U = 0 \mid R = 0) = 0.88$
$P(U = 1 \mid R = 0) = 0.12$

1. Initialize R and U with some values in the support
2. For *i* iterations
   1. **Obtain new $\boldsymbol{R}$ from $\boldsymbol{P(R \mid U)}$**
   2. Obtain new U from $P(U \mid R)$

Gibbs sampling - example

$P(R,U)$	Raining: 1	Raining: 0
Umbrella: 1	2	2
Umbrella: 0	0	2

$\boldsymbol{P(R = 1 \mid U = 1) = 0.6}$
$\boldsymbol{P(R = 0 \mid U = 1) = 0.4}$
$P(R = 0 \mid U = 0) = 0.93$
$P(R = 1 \mid U = 0) = 0.07$

$P(U = 1 \mid R = 1) = 0.75$
$P(U = 0 \mid R = 1) = 0.25$
$P(U = 0 \mid R = 0) = 0.88$
$P(U = 1 \mid R = 0) = 0.12$

1. Initialize R and U with some values in the support
2. For *i* iterations
   1. **Obtain new R from $P(R \mid U)$**
   2. Obtain new U from $P(U \mid R)$

Gibbs sampling - example

$P(R, U)$	Raining: 1	Raining: 0
Umbrella: 1	2	2
Umbrella: 0	0	2

$$P(R = 1 \mid U = 1) = 0.6$$
$$P(R = 0 \mid U = 1) = 0.4$$
$$P(R = 0 \mid U = 0) = 0.93$$
$$P(R = 1 \mid U = 0) = 0.07$$

$$\boldsymbol{P(U = 1 \mid R = 1) = 0.75}$$
$$\boldsymbol{P(U = 0 \mid R = 1) = 0.25}$$
$$P(U = 0 \mid R = 0) = 0.88$$
$$P(U = 1 \mid R = 0) = 0.12$$

1. Initialize R and U with some values in the support
2. For *i* iterations
   1. Obtain new R from $P(R \mid U)$
   2. **Obtain new U from $P(U \mid R)$**

Gibbs sampling - example

$P(R,U)$	Raining: 1	Raining: 0
Umbrella: 1	2	2
Umbrella: 0	1	2

$P(R = 1 \mid U = 1) = 0.6$
$P(R = 0 \mid U = 1) = 0.4$
$P(R = 0 \mid U = 0) = 0.93$
$P(R = 1 \mid U = 0) = 0.07$

$\boldsymbol{P(U = 1 \mid R = 1) = 0.75}$
$\boldsymbol{P(U = 0 \mid R = 1) = 0.25}$
$P(U = 0 \mid R = 0) = 0.88$
$P(U = 1 \mid R = 0) = 0.12$

1. Initialize R and U with some values in the support
2. For *i* iterations
    1. Obtain new R from $P(R \mid U)$
    2. **Obtain new U from $P(U \mid R)$**

Gibbs sampling - example

$P(R,U)$	Raining: 1	Raining: 0
Umbrella: 1	2	2
Umbrella: 0	1	2

$P(R = 1 \mid U = 1) = 0.6$
$P(R = 0 \mid U = 1) = 0.4$
$\boldsymbol{P(R = 0 \mid U = 0) = 0.93}$
$\boldsymbol{P(R = 1 \mid U = 0) = 0.07}$

$P(U = 1 \mid R = 1) = 0.75$
$P(U = 0 \mid R = 1) = 0.25$
$P(U = 0 \mid R = 0) = 0.88$
$P(U = 1 \mid R = 0) = 0.12$

1. Initialize R and U with some values in the support
2. For *i* iterations
    1. **Obtain new $\boldsymbol{R}$ from $\boldsymbol{P(R \mid U)}$**
    2. Obtain new U from $P(U \mid R)$

Gibbs sampling - example

$P(R,U)$	Raining: 1	Raining: 0
Umbrella: 1	2	2
Umbrella: 0	1	3

$P(R = 1 \mid U = 1) = 0.6$
$P(R = 0 \mid U = 1) = 0.4$
$\boldsymbol{P(R = 0 \mid U = 0) = 0.93}$
$\boldsymbol{P(R = 1 \mid U = 0) = 0.07}$

$P(U = 1 \mid R = 1) = 0.75$
$P(U = 0 \mid R = 1) = 0.25$
$P(U = 0 \mid R = 0) = 0.88$
$P(U = 1 \mid R = 0) = 0.12$

1. Initialize R and U with some values in the support
2. For *i* iterations
   1. **Obtain new $\boldsymbol{R}$ from $\boldsymbol{P(R \mid U)}$**
   2. Obtain new U from $P(U \mid R)$

Gibbs sampling - example

$P(R, U)$	Raining: 1	Raining: 0
Umbrella: 1	2	2
Umbrella: 0	1	3

$$P(R = 1 \mid U = 1) = 0.6$$
$$P(R = 0 \mid U = 1) = 0.4$$
$$P(R = 0 \mid U = 0) = 0.93$$
$$P(R = 1 \mid U = 0) = 0.07$$

$$P(U = 1 \mid R = 1) = 0.75$$
$$P(U = 0 \mid R = 1) = 0.25$$
$$\boldsymbol{P(U = 0 \mid R = 0) = 0.88}$$
$$\boldsymbol{P(U = 1 \mid R = 0) = 0.12}$$

1. Initialize R and U with some values in the support
2. For *i* iterations
    1. Obtain new R from $P(R \mid U)$
    2. **Obtain new U from $P(U \mid R)$**

Gibbs sampling - example

$P(R, U)$		Raining	
		1	0
Umbrella	1	2	2
	0	1	4

$$P(R = 1 \mid U = 1) = 0.6$$
$$P(R = 0 \mid U = 1) = 0.4$$
$$P(R = 0 \mid U = 0) = 0.93$$
$$P(R = 1 \mid U = 0) = 0.07$$

$$P(U = 1 \mid R = 1) = 0.75$$
$$P(U = 0 \mid R = 1) = 0.25$$
$$\boldsymbol{P(U = 0 \mid R = 0) = 0.88}$$
$$\boldsymbol{P(U = 1 \mid R = 0) = 0.12}$$

1. Initialize R and U with some values in the support
2. For *i* iterations
   1. Obtain new R from $P(R \mid U)$
   2. **Obtain new U from $P(U \mid R)$**

Gibbs sampling - example

$P(R,U)$		Raining 1	Raining 0
Umbrella	1	2	2
	0	1	4

$P(R,U)$		Raining 1	Raining 0
Umbrella	1	0.15	0.1
	0	0.05	0.7

- After 4 iterations, the sampled joint distribution is not close to the actual joint distribution.

Gibbs sampling - example

$P(R,U)$		Raining 1	Raining 0
Umbrella	1	152	109
	0	42	698

$P(R,U)$		Raining 1	Raining 0
Umbrella	1	0.15	0.1
	0	0.05	0.7

- After 4 iterations, the sampled joint distribution is not close to the actual joint distribution.
- But if you repeat this process long enough, it will approximate the joint distribution.

My work

- Arrange probability distributions in a way such that when they are sampled from, they generate a knowledge graph.
    - Conditional distributions are possible to calculate using **Bayes' rule**.

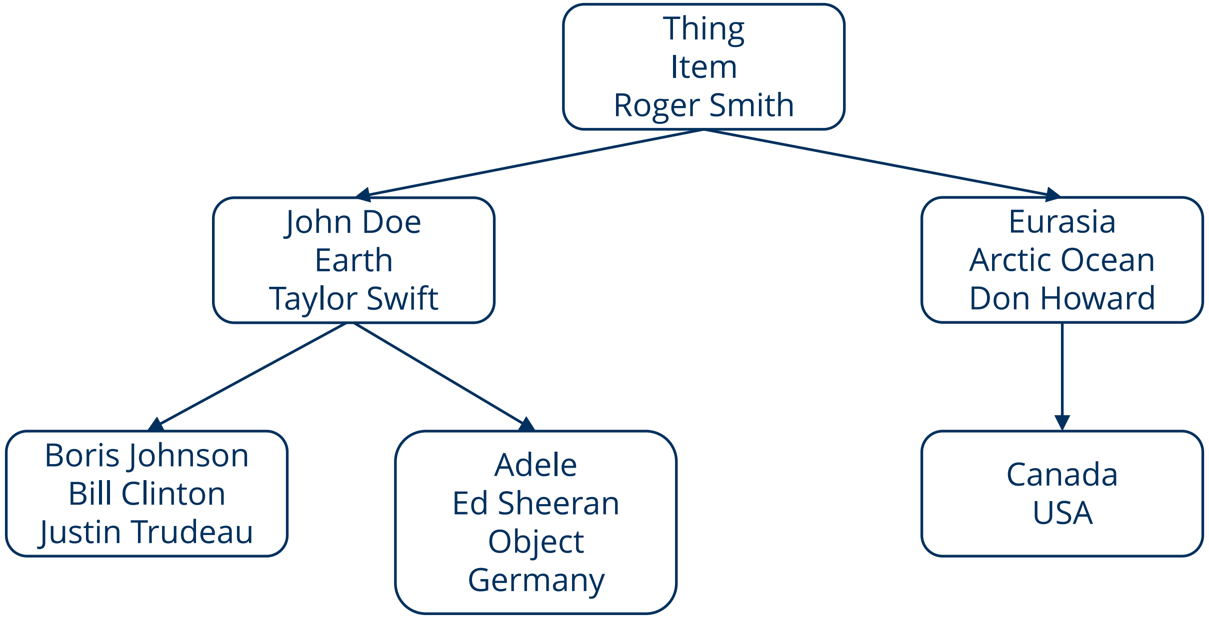
My work

- Arrange probability distributions in a way such that when they are sampled from, they generate a knowledge graph.
  - Conditional distributions are possible to calculate using **Bayes' rule**.
- We need to find the joint probability distribution of model parameters conditioned on the data.
  - $P(parameters \mid data)$

My work

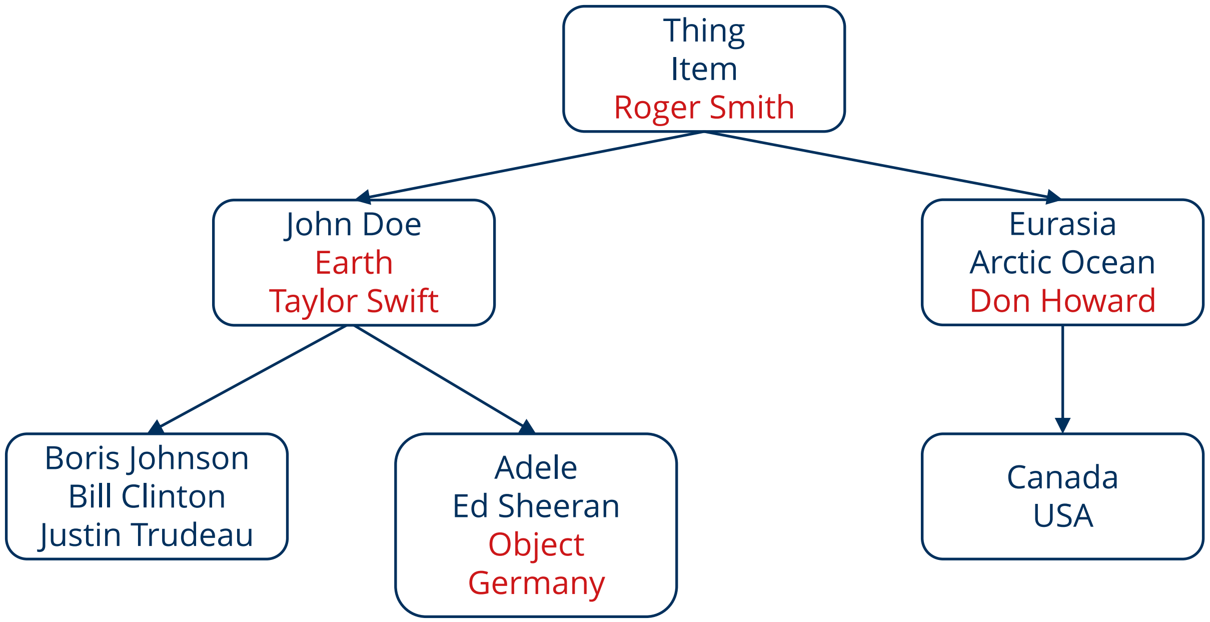
- Arrange probability distributions in a way such that when they are sampled from, they generate a knowledge graph.
  - Conditional distributions are possible to calculate using **Bayes' rule**.
- We need to find the joint probability distribution of model parameters and data.
  - $P(parameters \mid data)$
- We cannot find a closed form solution to $P(parameters \mid data)$ so we approximate it with **Gibbs sampling**.
  - In the model we will see, a tree gets generated, and two parameters get sampled: **paths** and **levels**.

My work - example



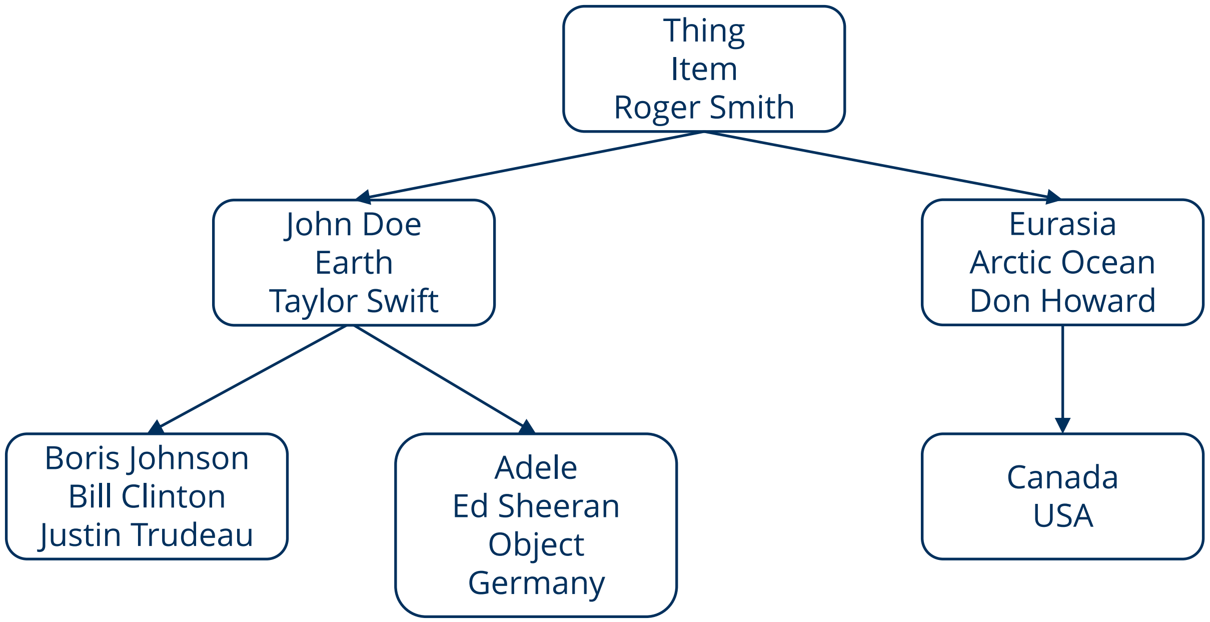
- Assume we have sampled the tree of knowledge graph entities above.

My work - example



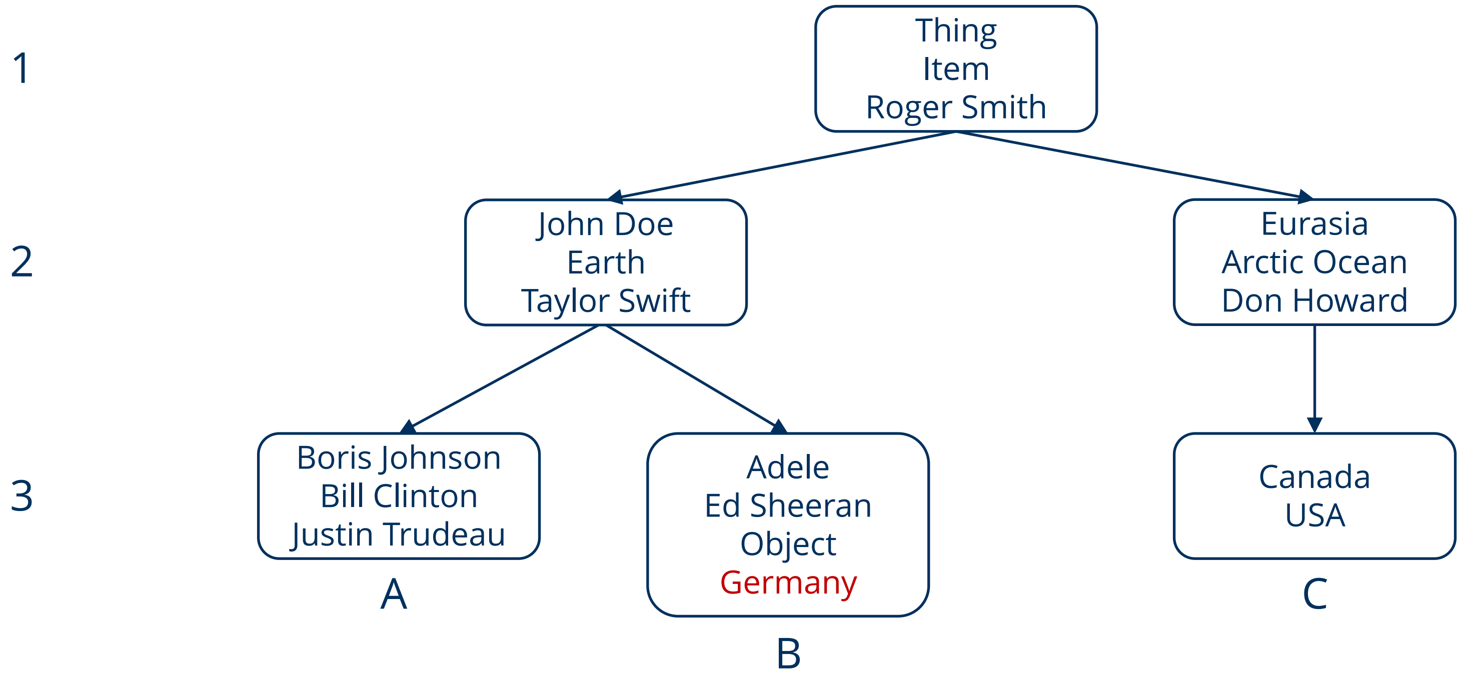
- Assume we have sampled the tree of knowledge graph entities above.
- Notice not all the entities are assigned a correct cluster.

My work - example



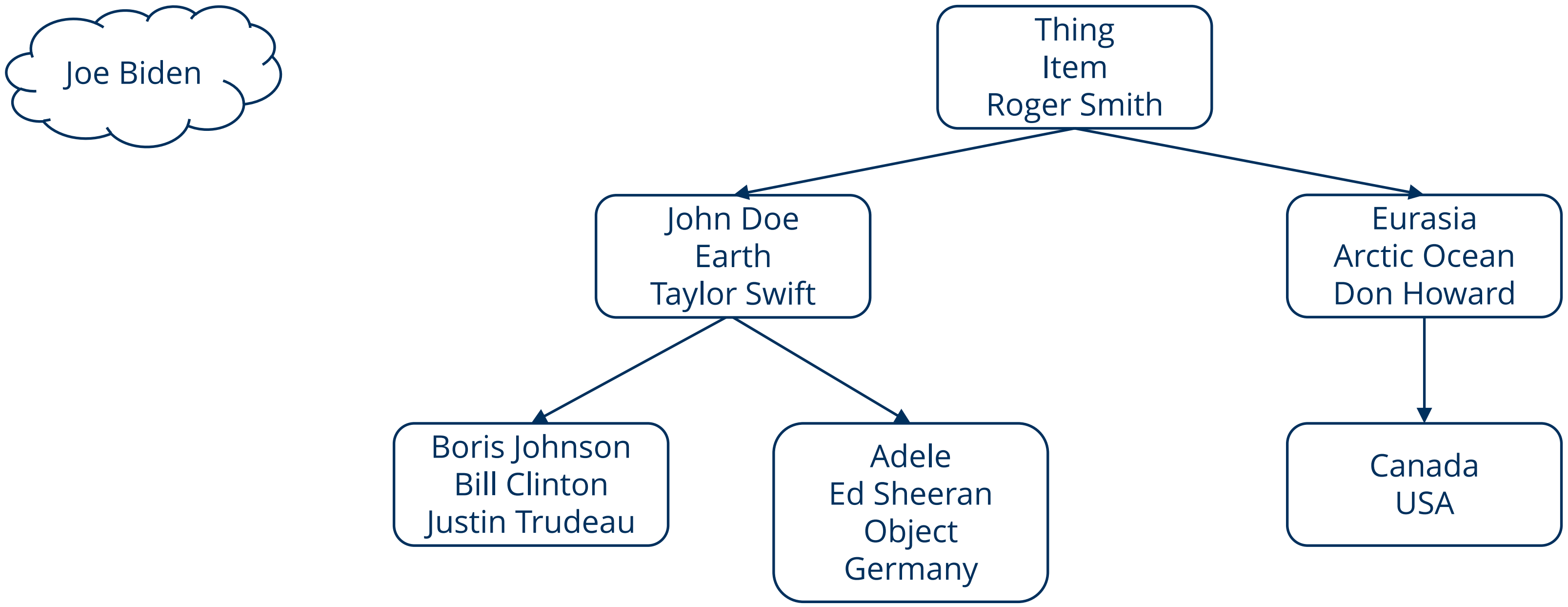
- The location of an entity on this tree is defined by its **path** and **level**.

My work - example



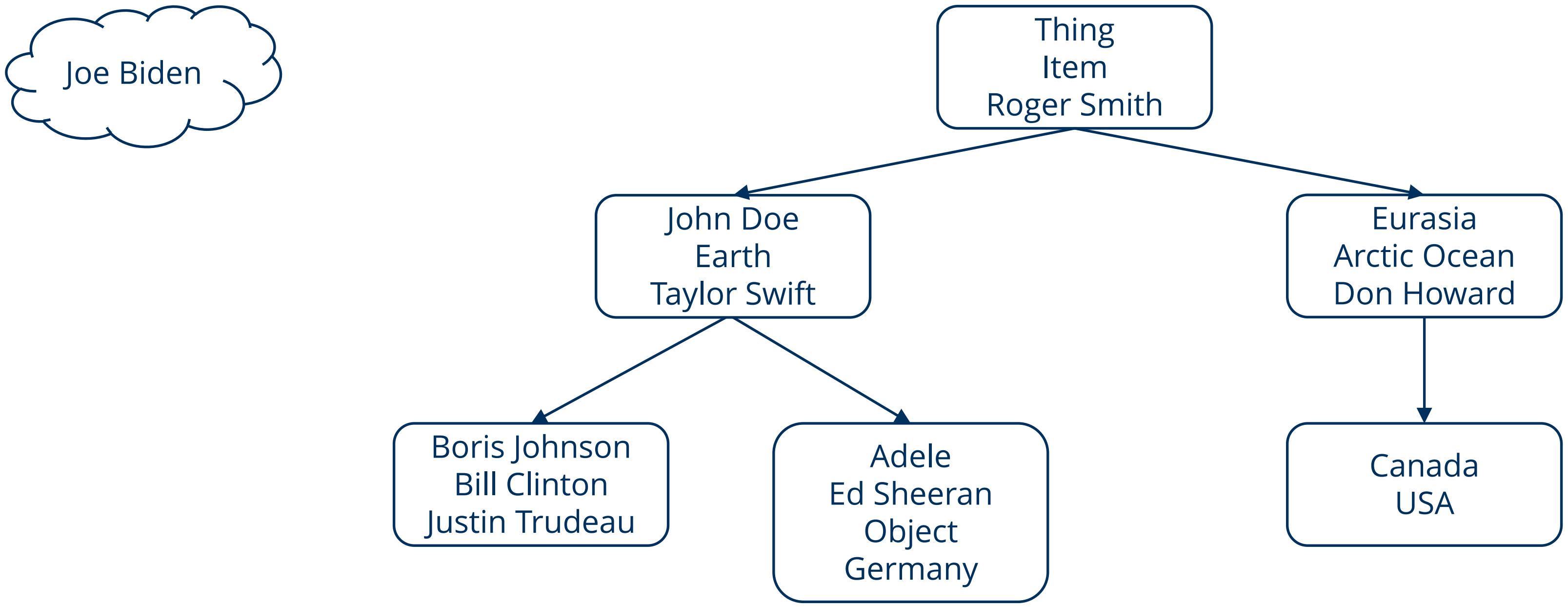
- The location of an entity on this tree is defined by its **path** and **level**.
- If we add labels, it becomes clearer.
  - For example, Germany is defined by path = B and level = 3.

My work - example



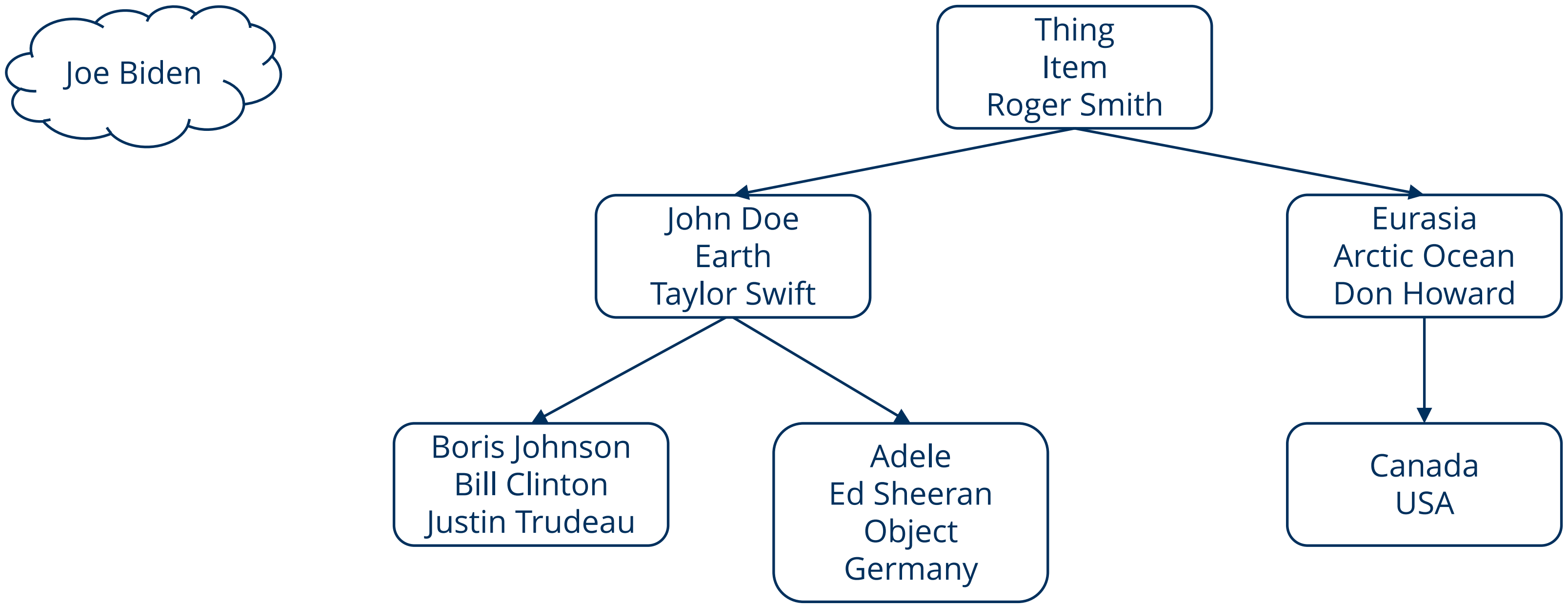
- Say Joe Biden is an entity we need to place on this tree.
  - We need to know its **path** and **level.**

My work - example



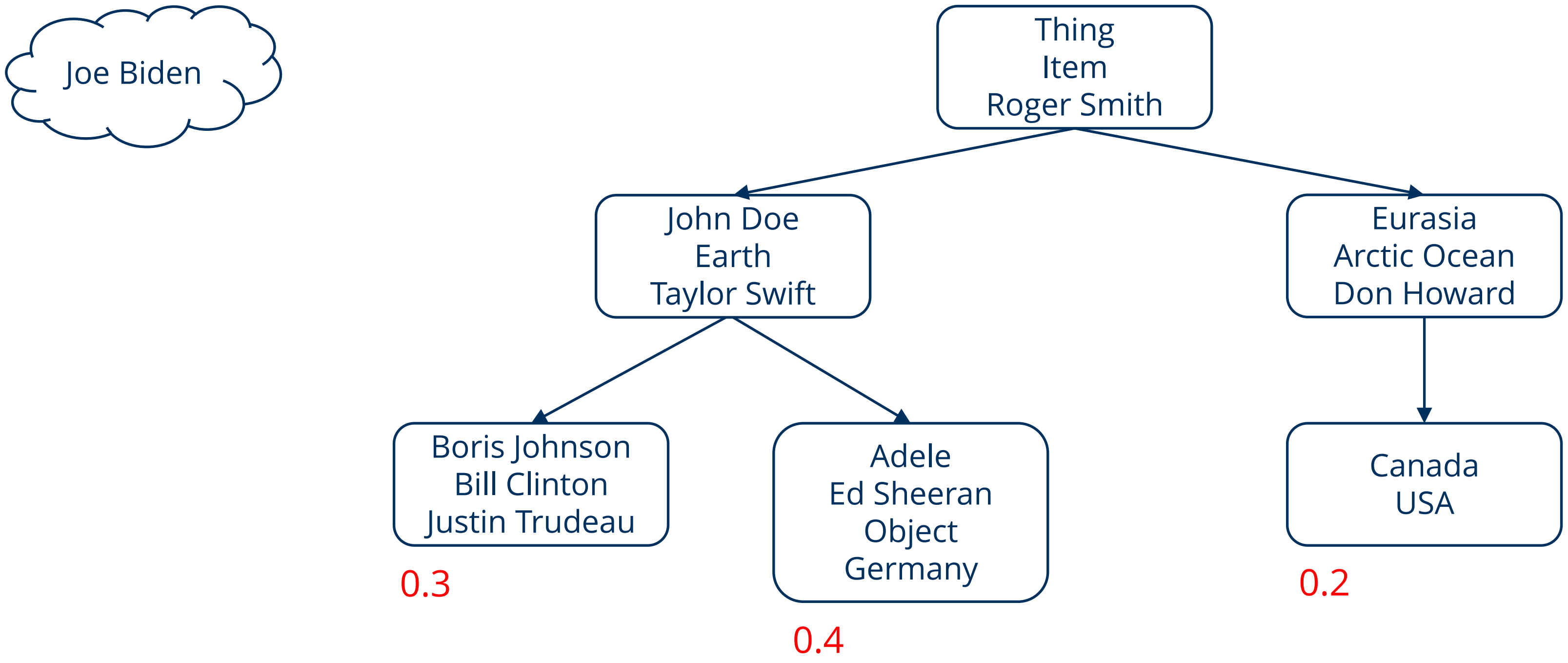
- Say Joe Biden is an entity we need to place on this tree.
  - We need to know its **path** and **level.**
  - These can be sampled from their conditional distributions!

My work - example



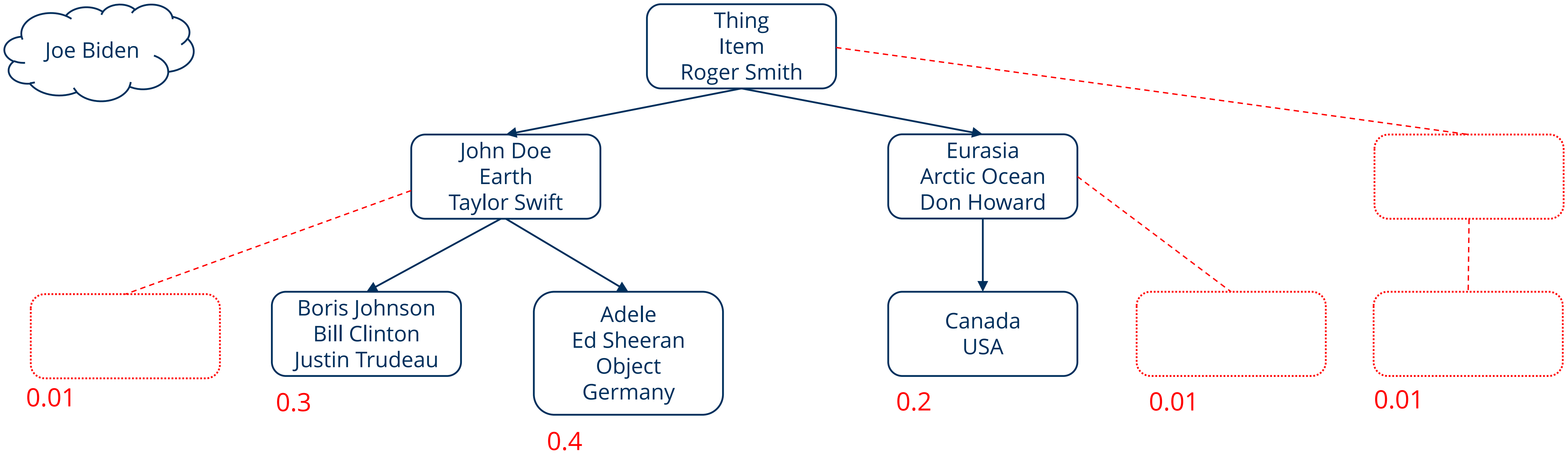
- $P(path \mid all\ other\ parameters, data) \propto P(data \mid path, all\ other\ parameters) P(path)$

My work - example



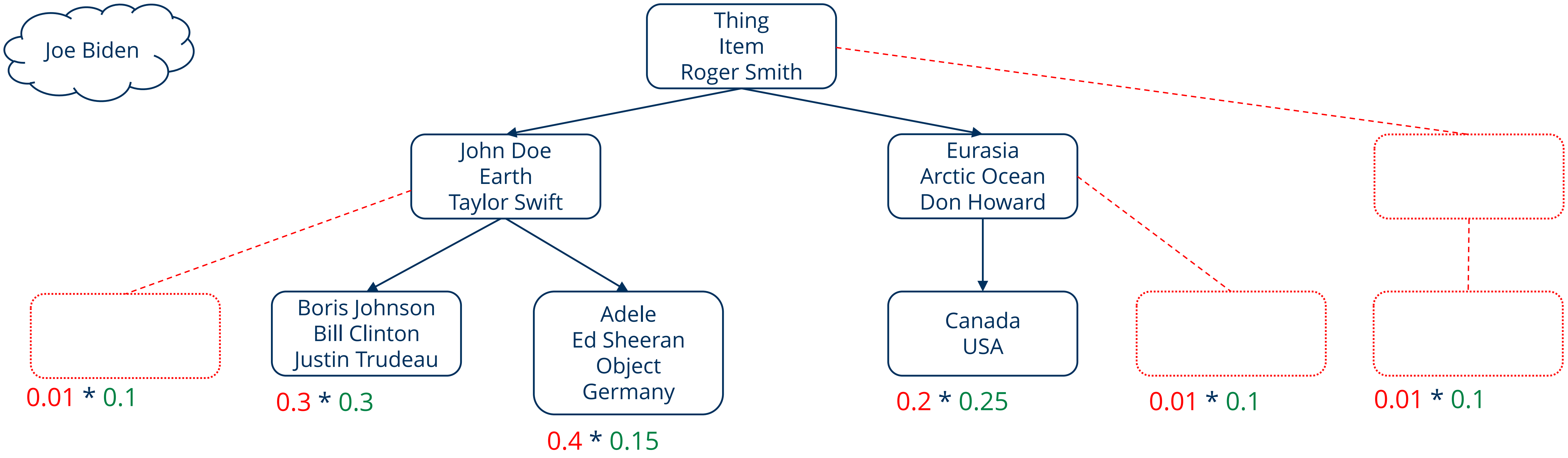
- $P(path \mid all\ other\ parameters, data) \propto {\color{green} P(data \mid path, all\ other\ parameters)}{\color{red} P(path)}$

My work - example



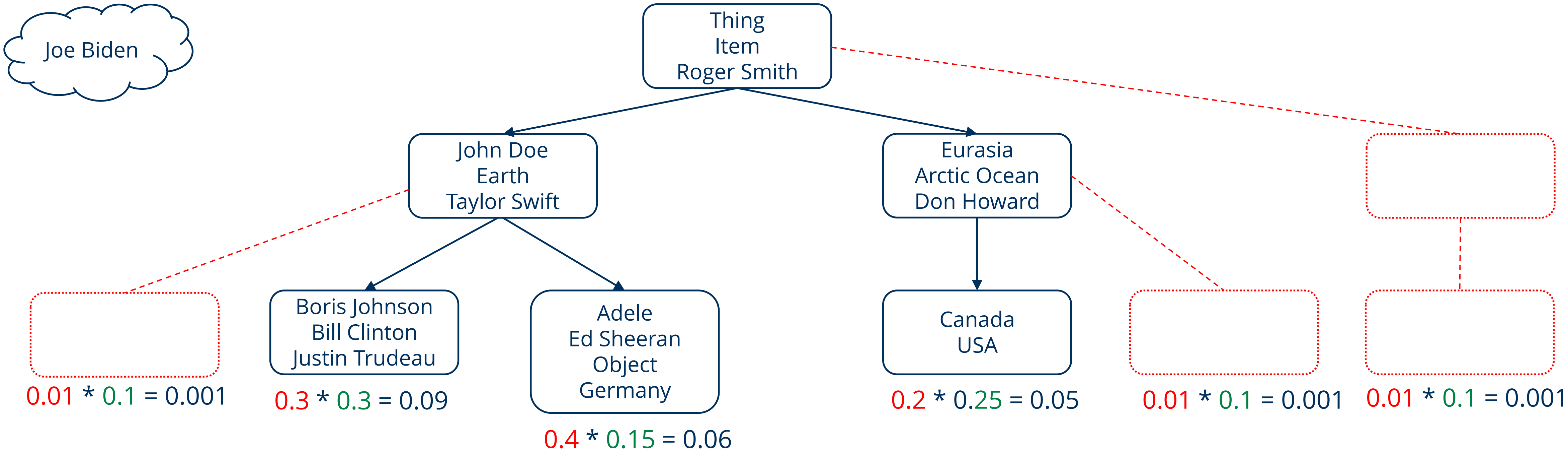
- $P(path \mid all\ other\ parameters, data) \propto P(data \mid path, all\ other\ parameters) P(path)$

My work - example



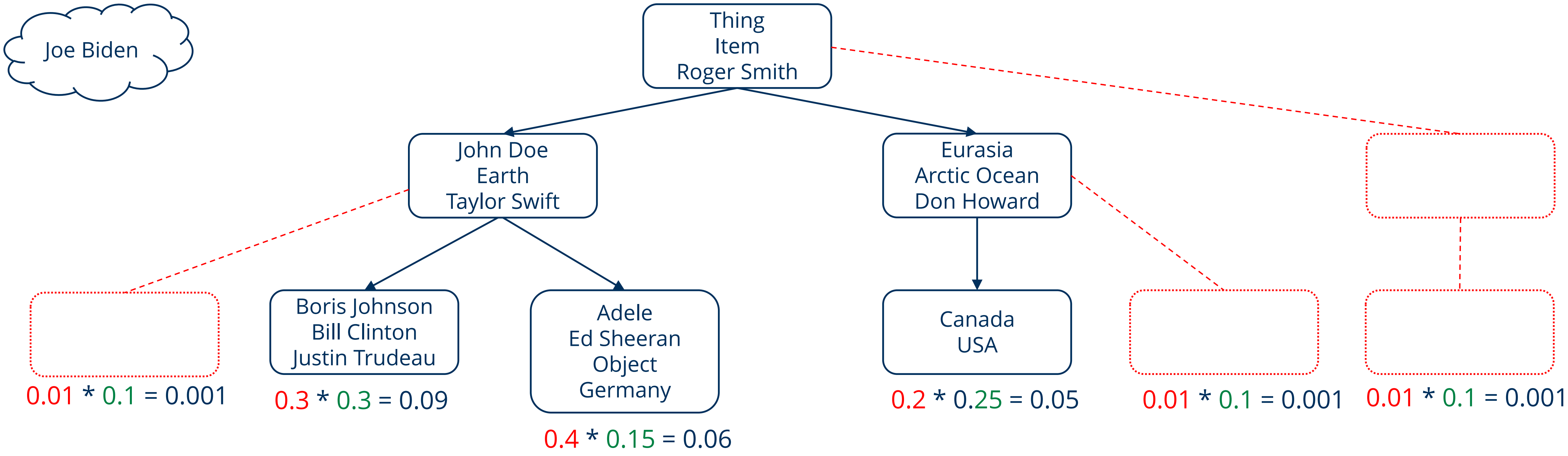
- $P(path \mid all\ other\ parameters, data) \propto P(data \mid path, all\ other\ parameters) P(path)$

My work - example



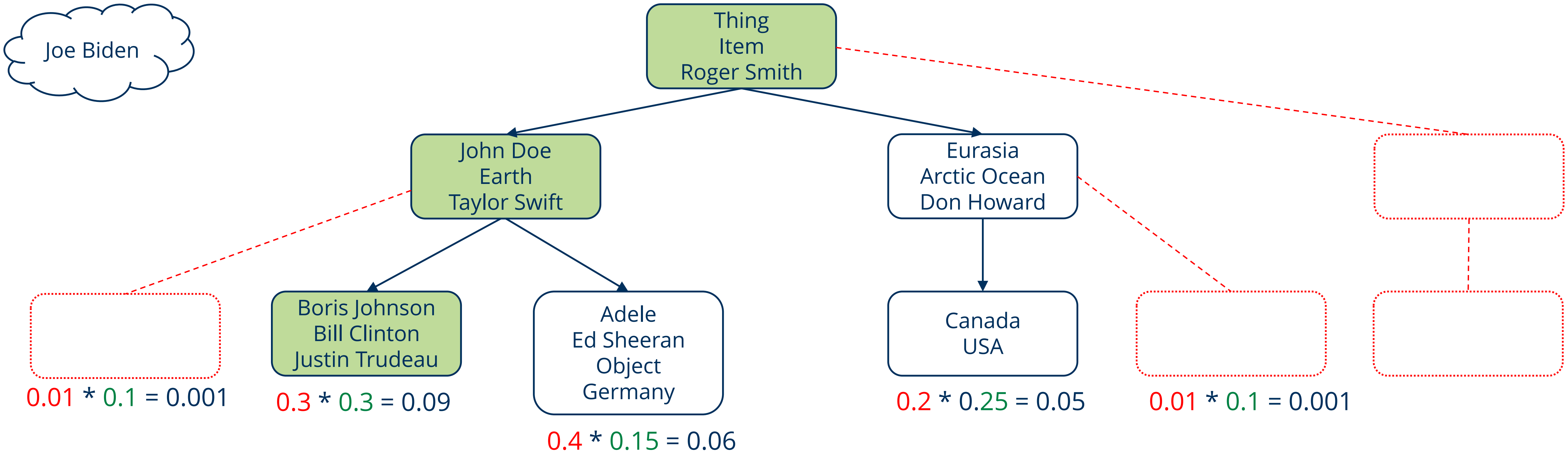
- $P(path \mid all\ other\ parameters, data) \propto P(data \mid path, all\ other\ parameters) P(path)$

My work - example



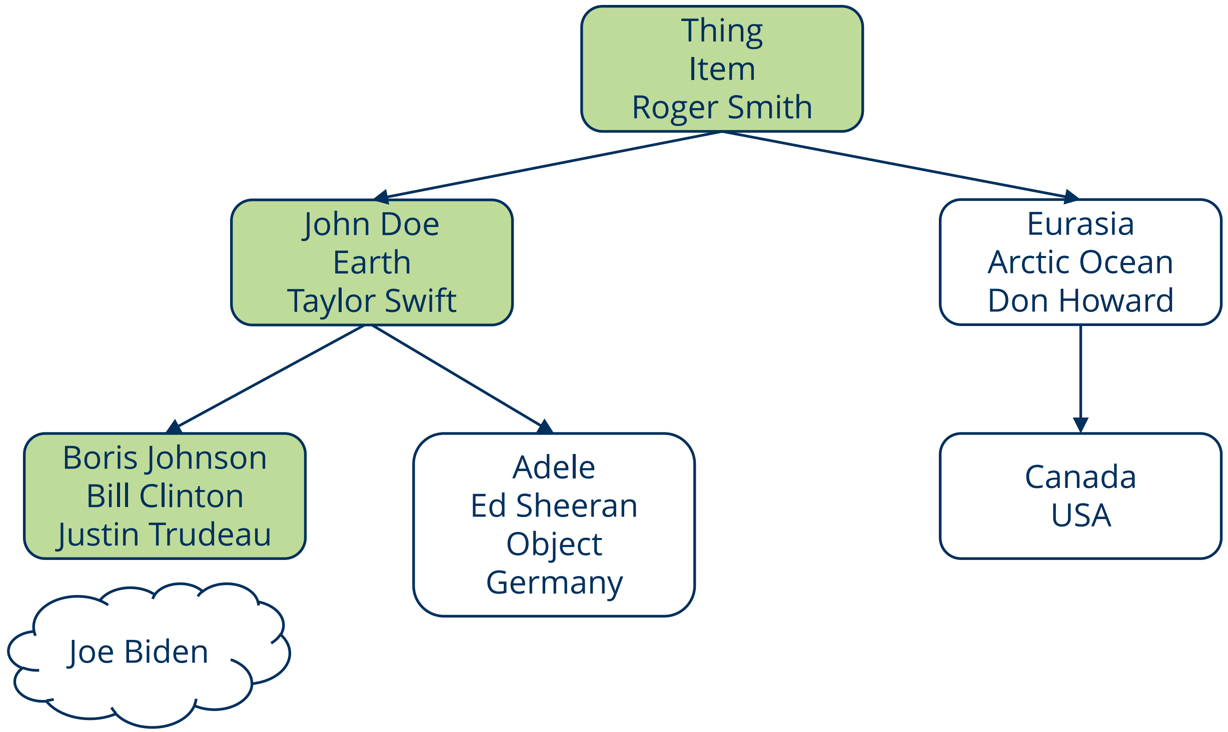
- $P(path \mid all\ other\ parameters, data) \propto P(data \mid path, all\ other\ parameters)P(path)$
- $P(path \mid all\ other\ parameters, data) \propto [0.001, 0.09, 0.06, 0.05, 0.001, 0.001]$
  - We can sample this!

My work - example



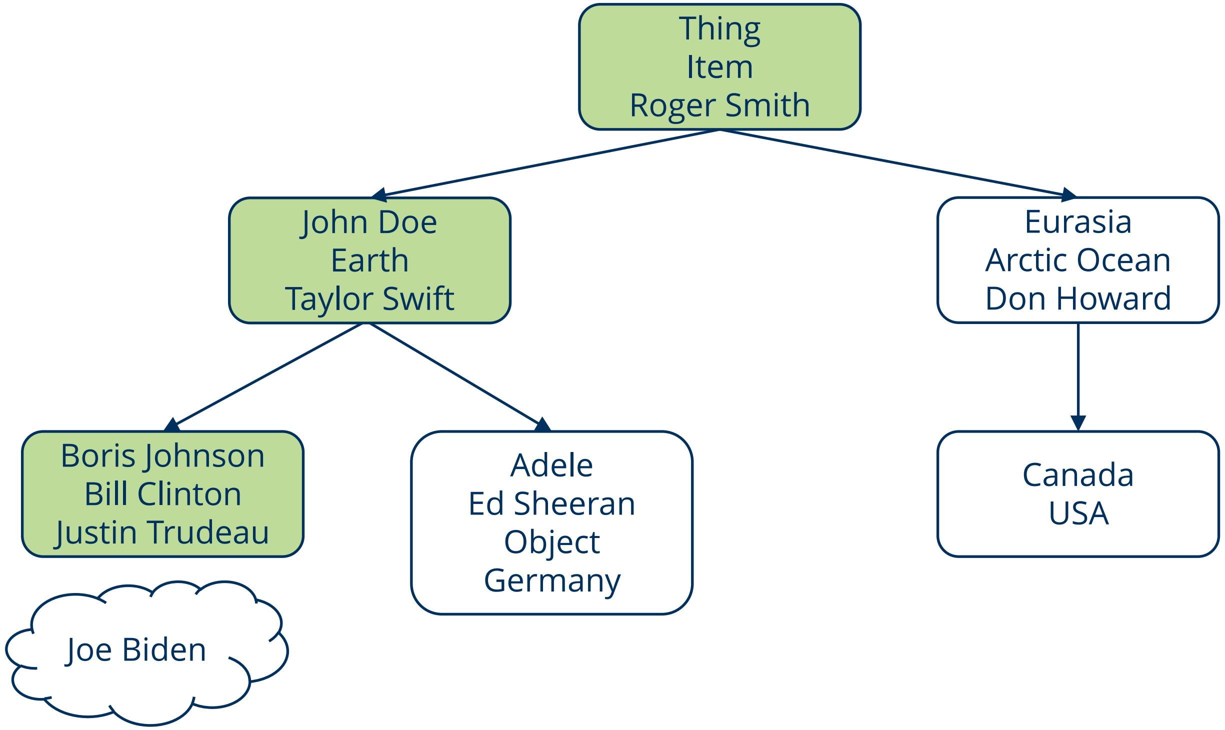
- Suppose we sample the path in green. We now know that Joe Biden will be placed in one of the three clusters on this path.

My work - example



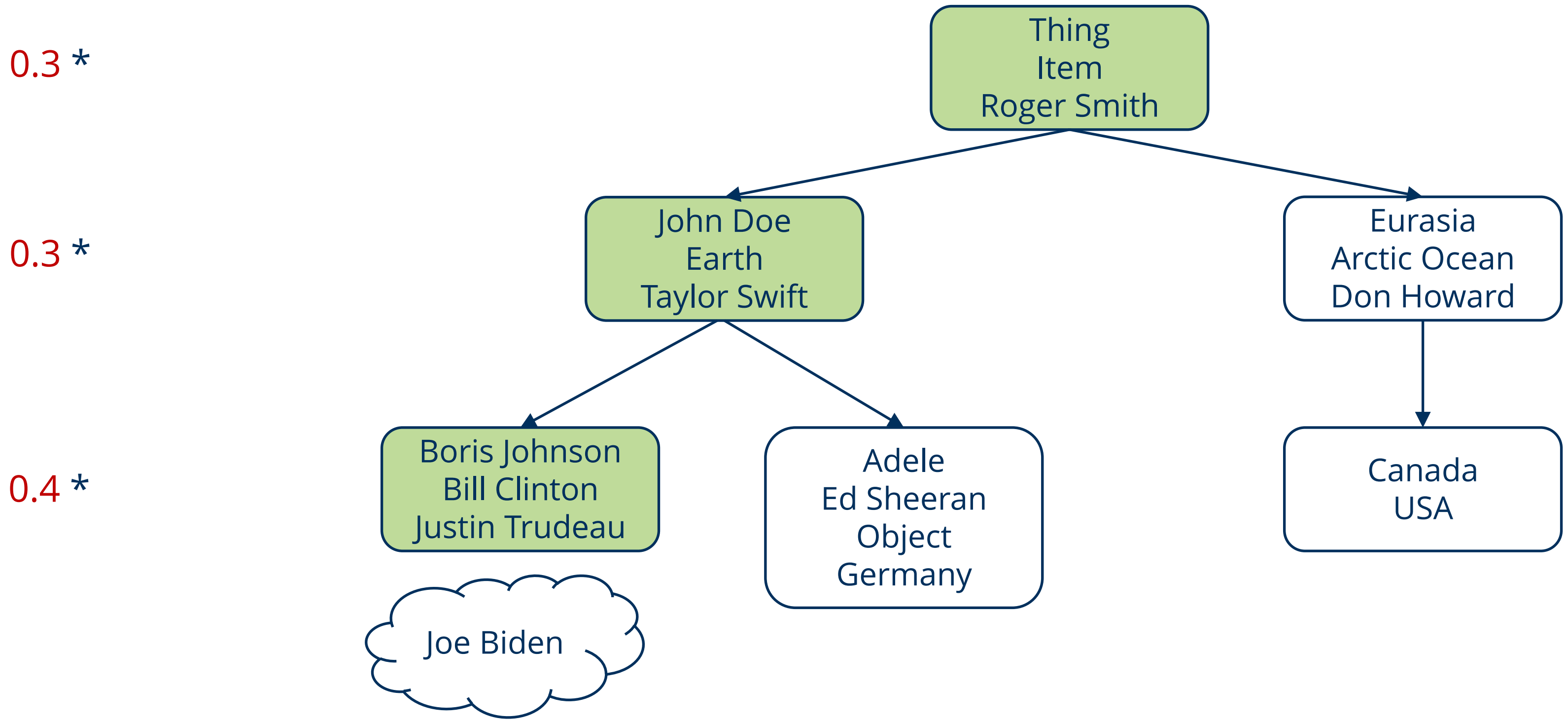
- Suppose we sample the path in green. We now know that Joe Biden will be placed in one of the three clusters on this path.
- Next, we sample the level to determine which cluster Joe Biden belongs to.

My work - example



- $P(level \mid all\ other\ parameters, data) \propto P(data \mid all\ other\ parameters, level)P(level)$

My work - example



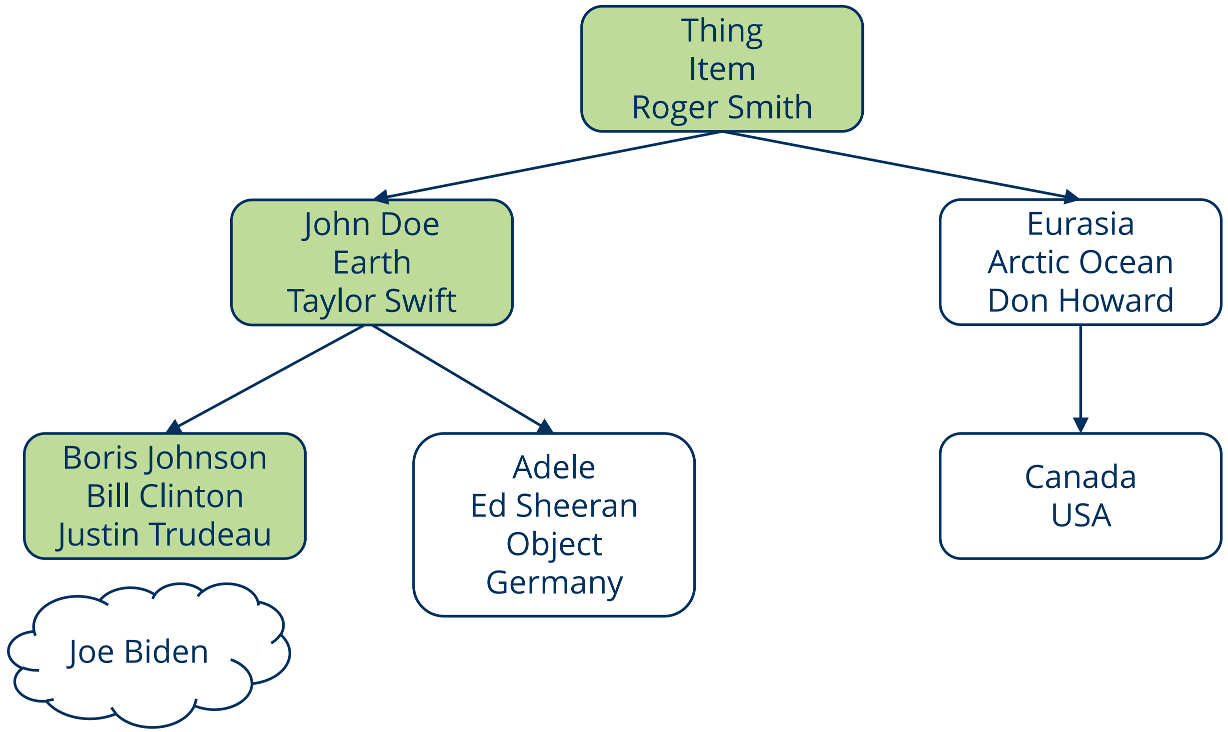
- $P(level \mid all\ other\ parameters, data) \propto P(data \mid all\ other\ parameters, level)P(level)$

My work - example

0.3 * 0.1 = 0.03

0.3 * 0.1 = 0.03

0.4 * 0.8 = 0.32



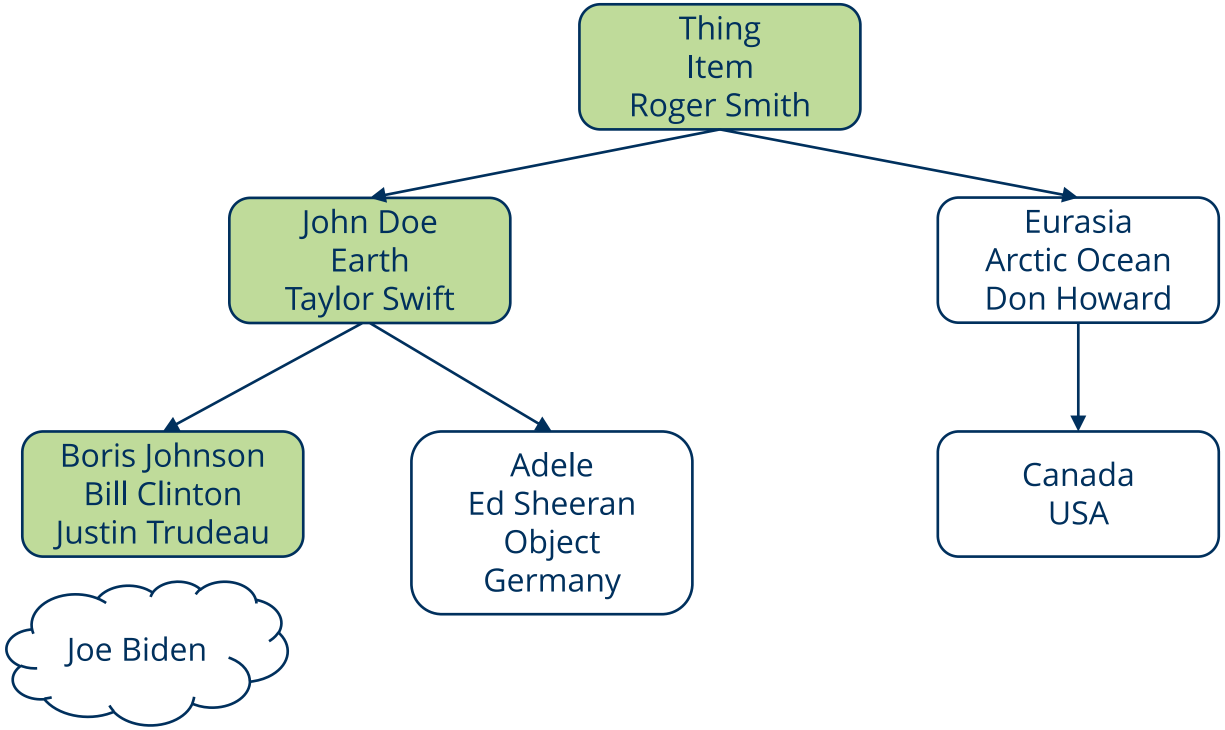
- $P(level \mid all\ other\ parameters, data) \propto P(data \mid all\ other\ parameters, level)P(level)$

My work - example

0.3 * 0.1 = 0.03

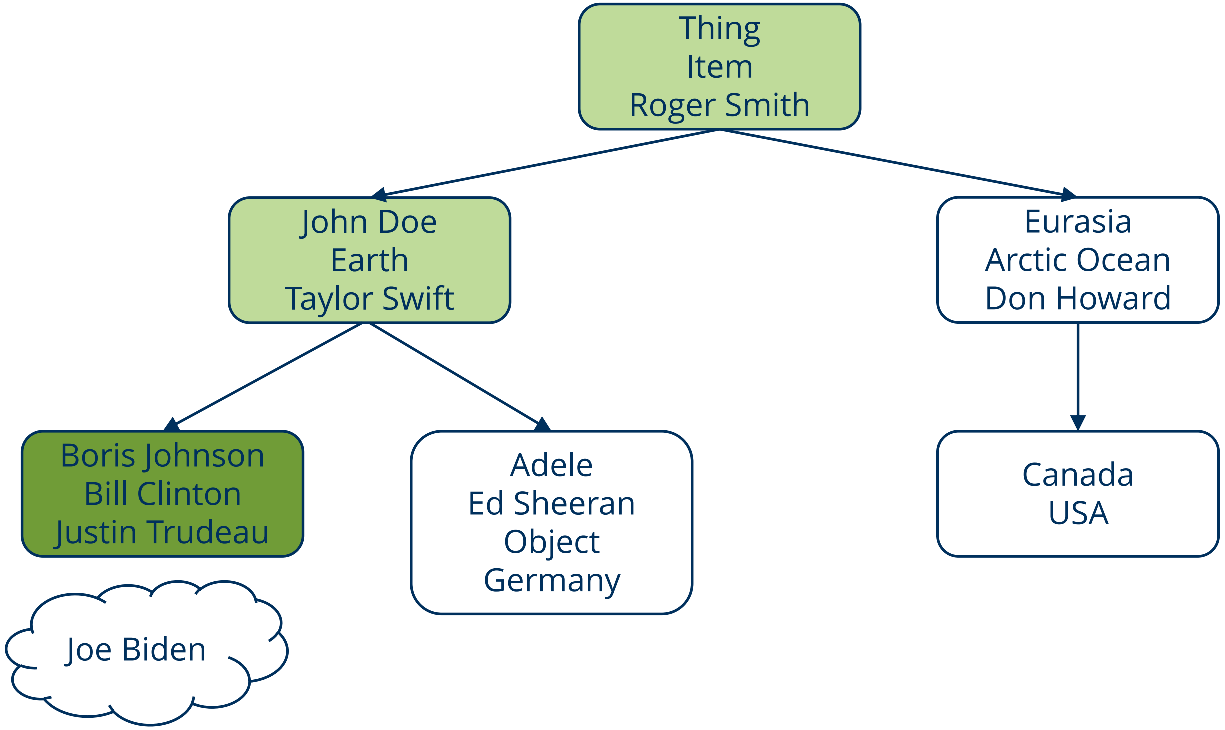
0.3 * 0.1 = 0.03

0.4 * 0.8 = 0.32



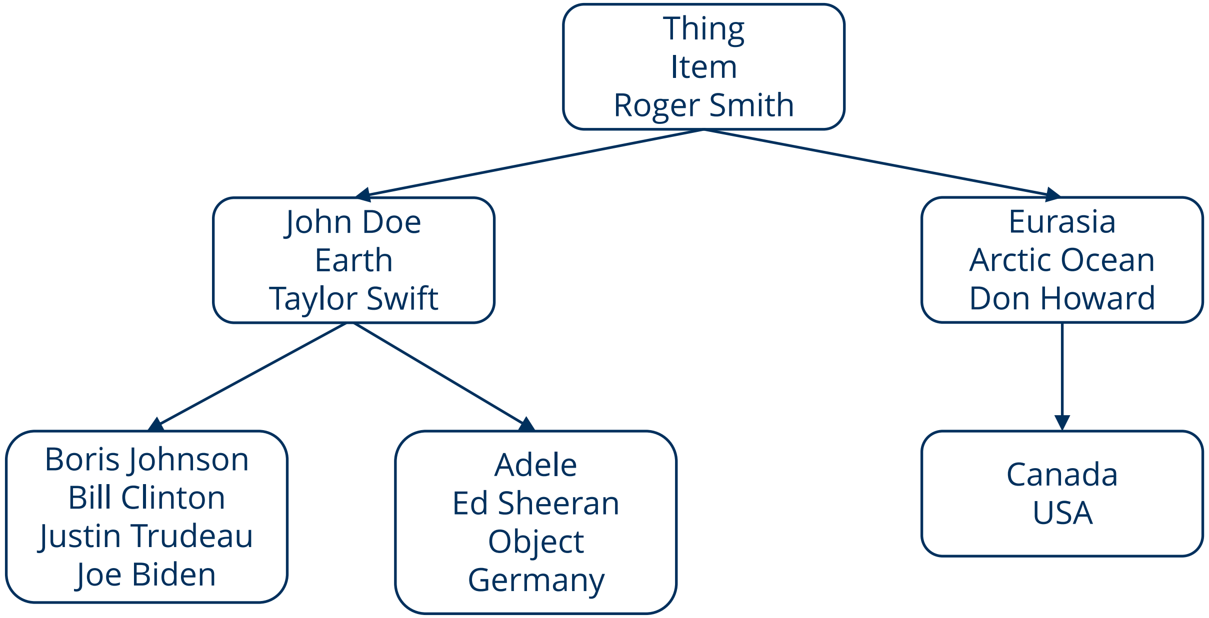
- $P(level \mid all\ other\ parameters, data) \propto P(data \mid all\ other\ parameters, level) P(level)$
- $P(level \mid all\ other\ parameters, data) \propto [0.03, 0.03, 0.32]$
  - Again, we can sample this.

My work - example



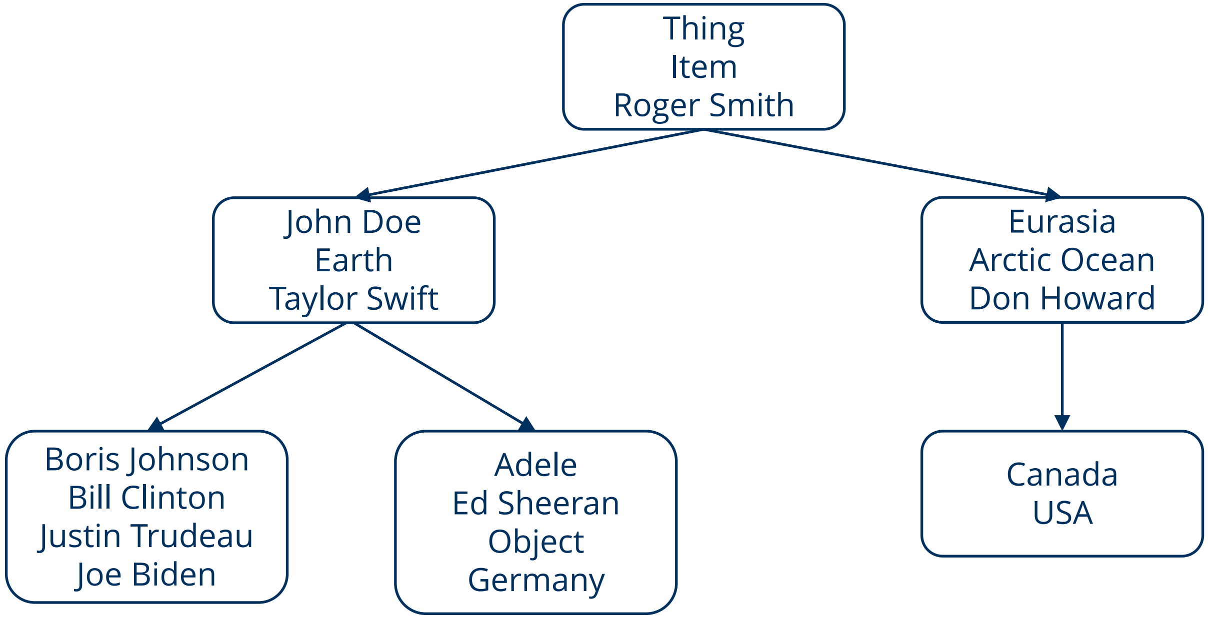
- Suppose we sample the bottom level. We can now put Joe Biden in the cluster at this level on the sampled path.

My work - example



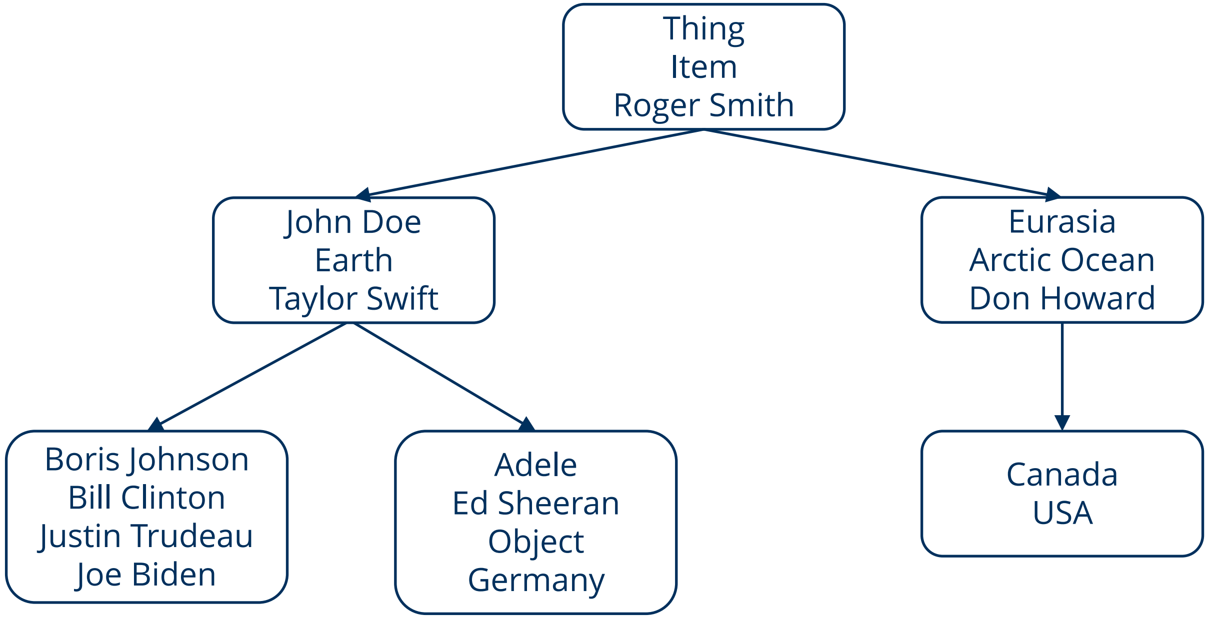
- Suppose we sample the bottom level. We can now put Joe Biden in the cluster at this level on the sampled path.
  - We have now sampled a cluster for an entity in the knowledge graph.

My work - example



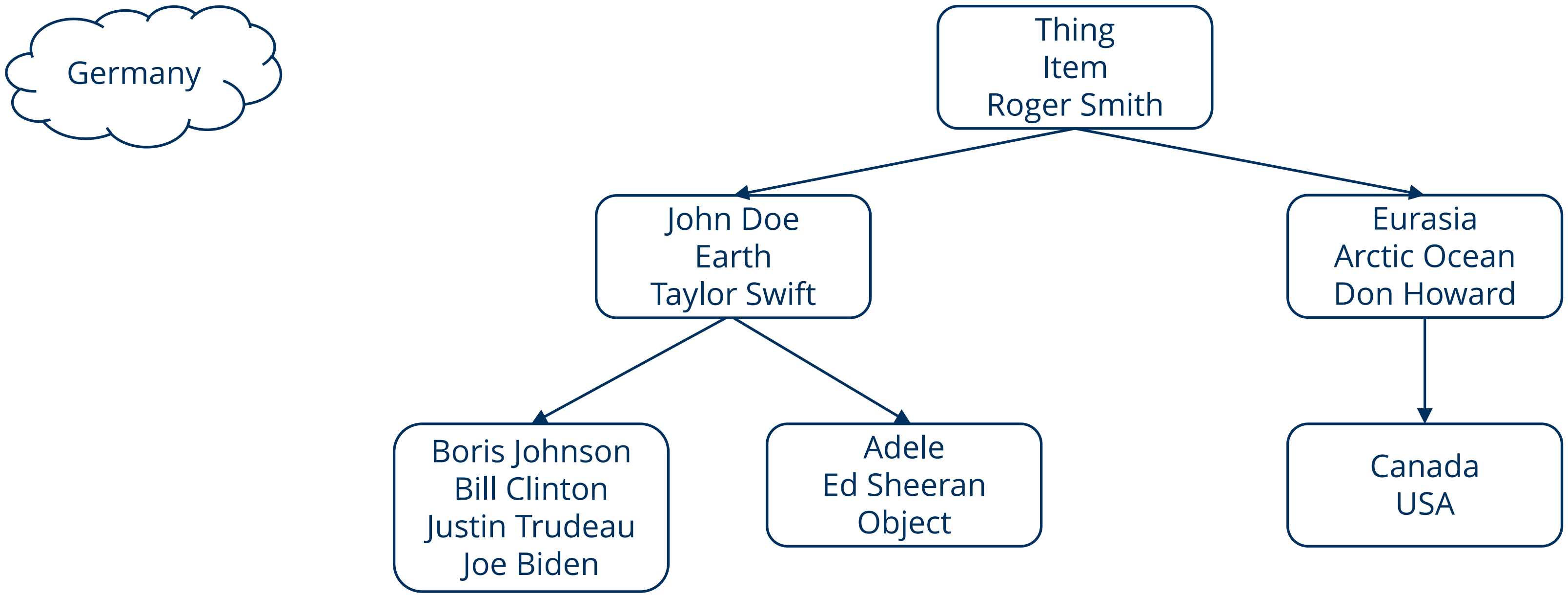
- But the tree is not correct yet… we must continue sampling.
  - Remove entity from tree and sample new path and level.

My work - example



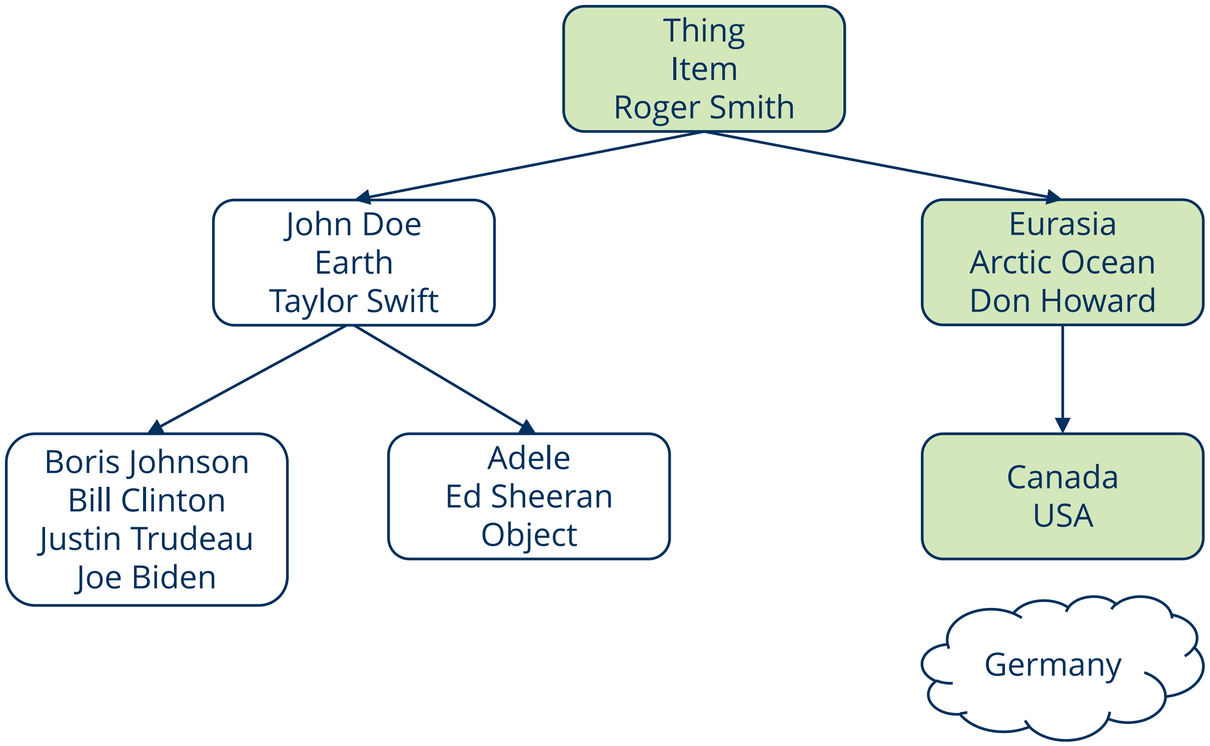
- Say the next entity we want to sample is Germany.

My work - example



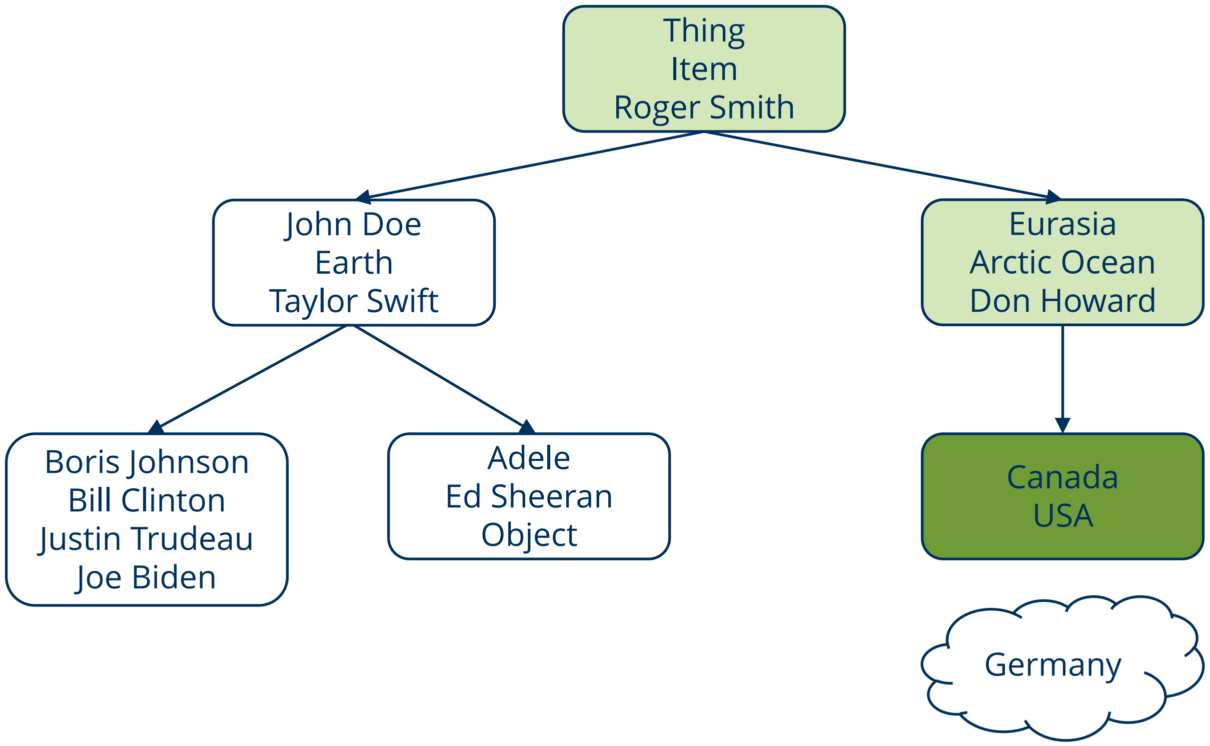
- Say the next entity we want to sample is Germany.
  - Remove Germany from tree.

My work - example



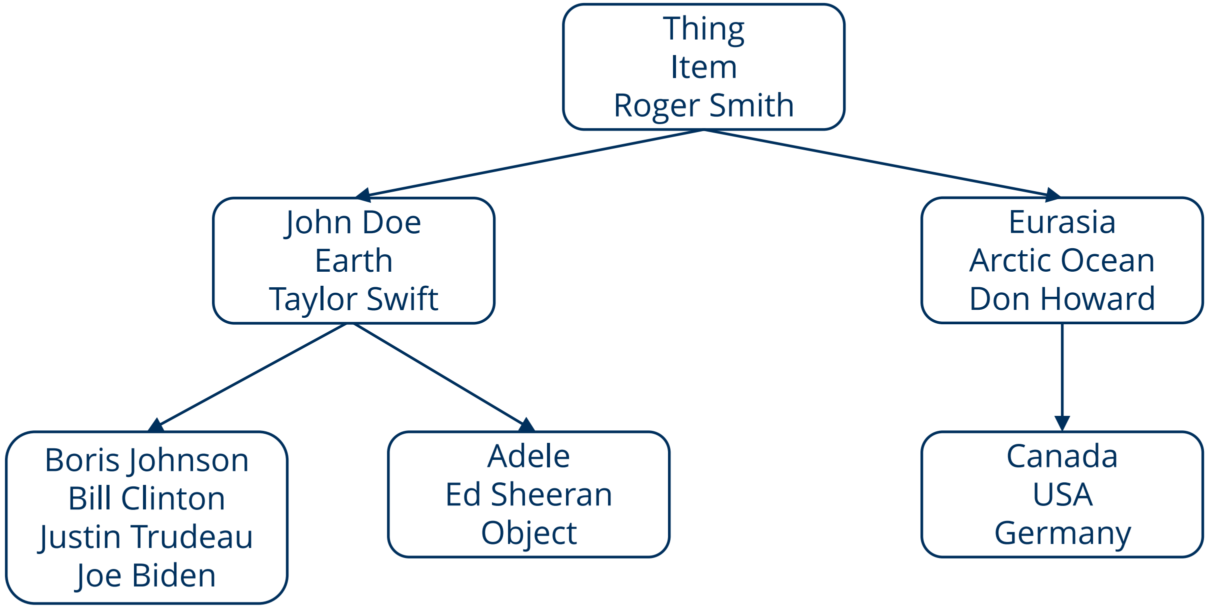
- Say the next entity we want to sample is Germany.
  - Remove Germany from tree.
  - Sample path

My work - example



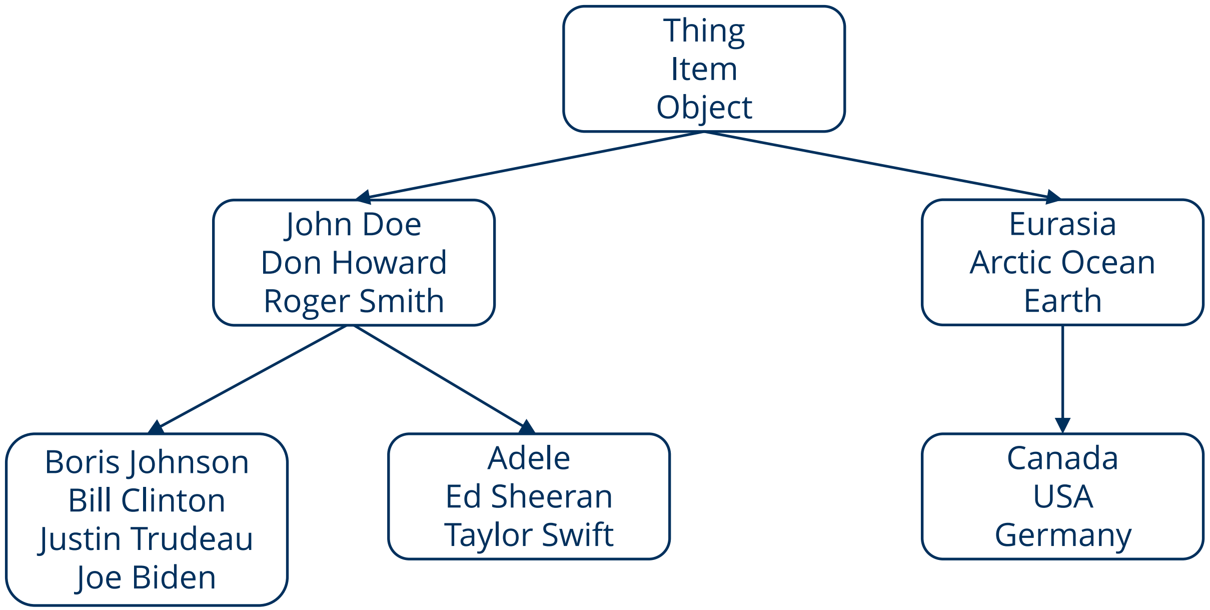
- Say the next entity we want to sample is Germany.
  - Remove Germany from tree.
  - Sample path and level.

My work - example



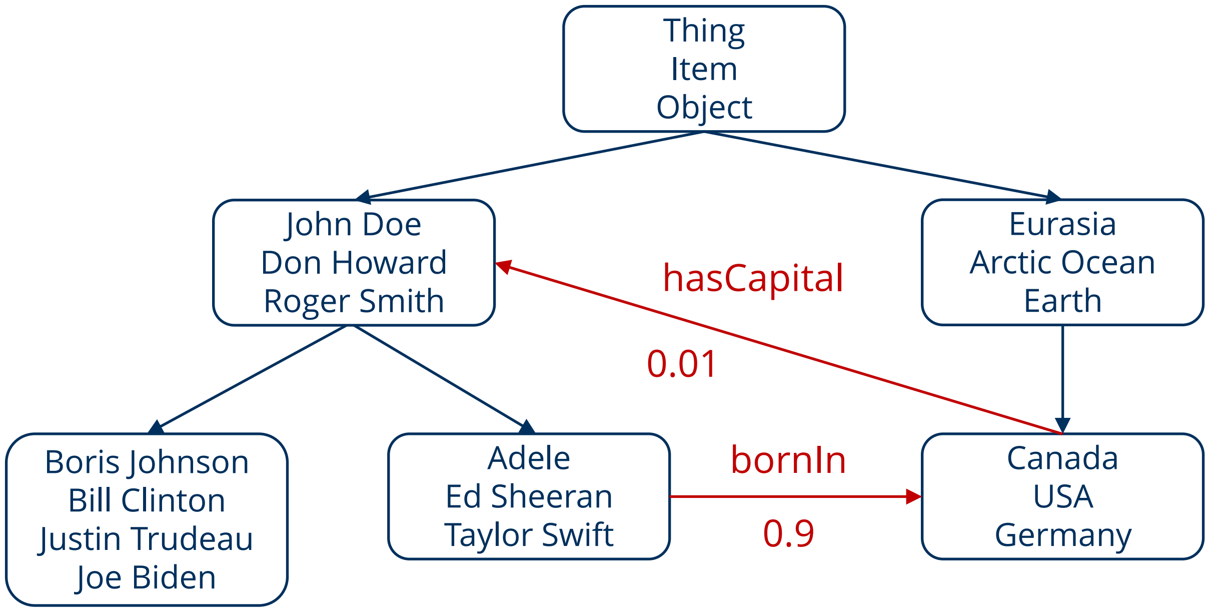
- We keep on sampling paths and levels for entities until some criteria is met.
  - This is how Gibbs sampling works in my model.

My work - example



- Ideally, we would want to get something like this.

My work - example



- Also, the captures the probabilities of interaction between cluster entities.
  - This is how the knowledge graph can get generated.

Summary

- Probabilistic methods can be used to model knowledge graphs.
  - Clustering
  - Topic modelling
  - Hierarchy induction
  - Etc.

Summary

- Probabilistic methods can be used to model knowledge graphs.
  - Clustering
  - Topic modelling
  - Hierarchy induction
  - Etc.
- The model I’m working on uses **Bayes’ rule** in conjunction with **Gibbs sampling** in its inference scheme.

Summary

- Probabilistic methods can be used to model knowledge graphs.
  - Clustering
  - Topic modelling
  - Hierarchy induction
  - Etc.
- The model I'm working on uses **Bayes' rule** in conjunction with **Gibbs sampling** in its inference scheme.
- Main problem is they're **too slow** (but solutions around the corner?).

Questions?